

# Observer variation in ophthalmology

R. A. GIBSON AND H. F. SANDERSON

*From the Bristol Eye Hospital, Bristol, and the Department of Community Health, London School of Hygiene and Tropical Medicine*

**SUMMARY** Data collected by clinicians and nurses from patients with clinically diagnosed lens opacities revealed that there was poor repeatability of the description of the position of lens opacities, statistically significant observer variation in determining visual acuity, and good repeatability as to what visual tasks the patient can perform. It is suggested that clinicians should give more weight to the visual handicap revealed by performance of visual tasks in making a decision for cataract surgery than at present.

Ophthalmologists mainly base their decision for or against advising cataract surgery on 3 criteria: visual acuity, position and severity of lens clouding, and the effect the lenticular opacities have on the patient's life style.<sup>1</sup> If the criteria are to be useful, methods of using them should be valid and repeatable. That is, they should measure what they purport to measure, and they should give the same answers on repeated request. However, measures that have a high degree of repeatability are not necessarily valid.

There is little reported information on repeatability of clinical methods of assessment in ophthalmology or of questionnaire data on visual disability.<sup>2</sup> It was therefore thought desirable, while carrying out a 2-centre study into the relative severity of cataract in groups of patients,<sup>3</sup> also to investigate the amount of variability which occurred in examination findings and questionnaire responses so that the repeatability of the currently employed methods might be assessed.

Disagreements in repeated observations on the same subjects by the same observer are termed 'within-observer error'. This may be due to inconsistency in technique or interpretation of results, and the variation tends to be random. Variation which occurs between different observers when they assess the same subjects is termed 'between-observer error', is usually due to differences in technique or interpretation, and tends to be systematic.

Two separate investigations of repeatability of clinical and interview results were carried out. The first indexed the collection of 54 separate items of information on clinical findings, personal data,

visual abilities, and past medical and ophthalmic history on 64 patients attending the Royal United Hospital, Bath, and the Bristol Eye Hospital. The second investigation was instituted to determine the repeatability of visual acuity measurements and was carried out in the outpatient department of the Bristol Eye Hospital.

## Patients and methods

### PATIENTS AND INTERVIEWS

Three hundred patients over 65 years old attending the outpatient department of the Royal United Hospital, Bath, and the Bristol Eye Hospital, who were noted on examination to have lens opacities which reduced the visual acuity to 6/9 or less, were interviewed and had an ophthalmic examination. All interviews were conducted by H.F.S., who also abstracted clinical data from the hospital notes and supplemented them by direct questioning of the clinician who had examined the patient. Patients who were fit enough to travel by car and who lived close to one or other of the 2 hospitals were invited to attend for re-examination; of the 300, 64 agreed to do so. Statistical examination of the data obtained about these 64 patients showed that this method of selection did not introduce bias.<sup>3</sup>

At the return visit all the 64 patients had an ophthalmic examination performed by R.A.G. with equipment similar to that used at the first examination. Interview was carried out by H.F.S. Re-examination and interview were performed without reference to the previous records in order to minimise possible bias introduced by knowledge of the original findings. The time that elapsed between first and second examination varied between three weeks and two months. Records of the first and

Correspondence to R. A. Gibson FRCS Ed., Bristol Eye Hospital, Lower Maudlin Street, Bristol BS1 2LX.

second examinations were compared and the percentage of agreements for each variable calculated to give a crude agreement rate.

#### VISUAL ACUITY MEASUREMENTS

Two nurses employed in the outpatient department of the Bristol Eye Hospital who regularly carried out visual acuity measurements agreed to participate in this study. Patients attending the clinic for any reason were tested first by one nurse and then by the other. The two nurses examined distance vision at 6 metres with illuminated Snellen charts. A different Snellen face was used in each of the 2 testing stations. Near visual acuity was tested with Faculty of Ophthalmologists approved test types (Curry and Paxton).

So that any differences in the results of visual acuity testing such as may occur due to differences between nurses, between stations, or between first and second examinations could be assessed, a block design was used.<sup>4</sup> Patients were allocated alternately to nurse 1 or nurse 2, and after each nurse had examined 6 patients the nurses changed station. This ensured that equal numbers of patients were examined by each nurse and in both stations. The distribution of patients between nurses and stations is shown in Table 1.

In comparing the first examination with the second examination differences in findings were designated positive if a smaller test type was read on the second examination and negative if a larger test type was read on the second examination. The same convention was used to compare nurse 1 and nurse 2 and station 1 and station 2.

## Results

#### CLINICAL OBSERVATIONS

Eighteen different clinical observations were made twice on each of the 64 patients, giving in total 1152 pairs of observations. Of these pairs 207 did not agree, resulting in an overall crude agreement rate of 82%.

Table 2 shows the findings of the clinical examination.

Table 1 *Repeatability of visual acuity measurements, number of patients by station, nurse, and order of examination*

	Nurse 1		Nurse 2		Total
	Station 1	Station 2	Station 1	Station 2	
Examination 1	12	9	9	10	40
Examination 2	10	9	9	12	40
Total	22	18	18	22	80

Total number of patients = 40 = 80 eyes.

Table 2 *Examination—re-examination agreement rates for clinical findings*

Variable	Descriptive categories	Agreement rate (%)
Amblyopia	Present/absent	100
Glaucoma	Present/absent	96.8
Myopia	Present/absent	95.3
Lens	Normal/cataract/lens absent	93.8
Cataract	Mature/immature	91
Vitreous	Normal/abnormal	90
Cornea	Normal/central abnormality/peripheral abnormality	88
Retina	Normal/abnormal	79
Position of opacity	Nuclear/cortical/mixed	48

Number of patients = 64 (Bath 22, Bristol 42).

Table 3 *Agreement rates for interview data*

Variable	Agreement rate (%)
Ability to see window during the day	100
Ability to move about the house without help	100
Ability to recognise friends close up	96.8
Ability to watch television	96.8
Ability to read headlines in newspaper	96.8
Ability to do household cleaning and cookery	96.8
Ability to drive	93.7
Ability to go shopping	87.5
Ability to recognise friends across the street	84.4
Ability to cross streets unaided	82.5
Ability to recognise friends across the room	79.7
Ability to read small print in newspaper	76.5

Number of patients = 64 (Bath 22, Bristol 22).

nation. There was good agreement on the presence or absence of amblyopia, glaucoma, and myopia. The agreement on the presence of lens opacities and vitreous and corneal findings was less good. Approximately 1 in 10 was not agreed upon. Agreement rates for the diagnosis of an abnormal retina and in particular the description of position of lens opacity were much lower.

#### INTERVIEW DATA

At the interview each patient was asked to answer 35 questions, so that a total of 2,240 pairs of responses were recorded. 158 answers were not consistent in the 2 interviews, giving an overall crude agreement rate of 92.9%.

Table 3 shows the agreement rates for the interview data. Data on visual abilities such as reading and recognising friends had a higher agreement rate than the description of the position of lens opacity.

## VISUAL ACUITY RESULTS

Disagreements between nurses over the results of distance and near visual acuity testing were spread over the whole range of visual acuities.

Fig. 1 shows the distribution of the differences of distance visual acuity measurements on individual eyes when the results of nurse 1 are compared with those of nurse 2 (between nurses), the results of station 1 with those of station 2 (between stations), and the results of examination 1 with those of examination 2 (between examinations). Approximately symmetrical distributions for these differences were found. These distributions were tested for skewedness by Wilcoxon's signed rank test,<sup>4</sup> but in none of these is the distribution skewed more than might be expected by chance alone. Had the study been made with larger numbers of patients, a statistically significant systematic bias might have been revealed.

Fig. 2 also shows the distribution of the differences in near visual acuity tests, between nurses, stations, and examinations. The Wilcoxon's signed rank tests show that the distribution of the differences between nurses and between examinations are statistically significantly skewed both for the indi-

vidual eyes and for both eyes of the same patient ( $P = <0.01$  and  $<0.05$  respectively). This suggests that the second nurse consistently recorded patients as being able to read smaller print than the first nurse. Patients also read smaller print on the second examination. There was no significant skewing of the distribution of differences between stations.

## Discussion

A search of the literature on the epidemiology of ophthalmic conditions, and in particular cataract, was made to try and find reliable methods of examination and questionnaires of visual disability. Few studies have recognised that variations in clinical assessments can occur, either between different observers or between the same observer on different occasions, despite the fact that epidemiologists have been calling attention to these problems for more than 30 years.<sup>5</sup> Notable exceptions to this are the studies on glaucoma by Hollows and Graham<sup>6</sup> and the carefully carried out survey of ophthalmic conditions in the Framingham eye survey.<sup>2</sup> However, studies on the prevalence of abnormal findings in the eye are still published without any reference to the repeatability of the methods used.

The results reported here suggest that prevalence estimates based on clinical examination might vary by as much as 20% between different observers, and that this variation might be much higher for specific types of retinal pathology. The variation encountered in the description of position of lens opacity was even greater.

It is unlikely that this poor repeatability was due to special features of these patients. In addition, these patients were comparable with an outpatient sample from Coventry described by Brennan and Knox.<sup>7</sup> Poor repeatability of some clinical and interview data is therefore unlikely to be purely a local problem.

The repeatability of distance visual acuity tests was very low. The results of these tests showed that only a third of patients were recorded as having the same distance visual acuity on 2 occasions. Even when the results were analysed for individual eyes, fewer than half the observations on each eye gave the same results on the second test. The majority of the variation was only 1 line, but 13% of measurements on eyes differed by 2 lines or more. Near visual acuity measurements had a similar degree of variability; only a third of patients were recorded as having the same near visual acuity on both occasions. However, the range of variation was greater than for distance visual acuity measurements; 20% of all eyes were recorded as having a near visual

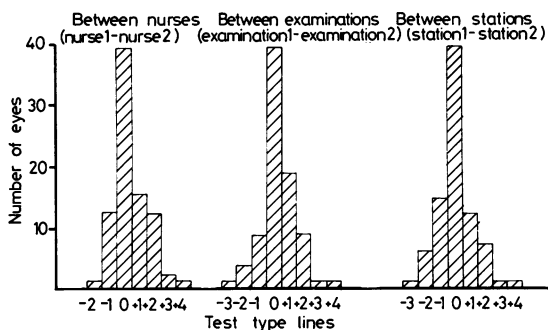


Fig. 1 Frequency distributions of test-retest differences in distance visual acuity testing

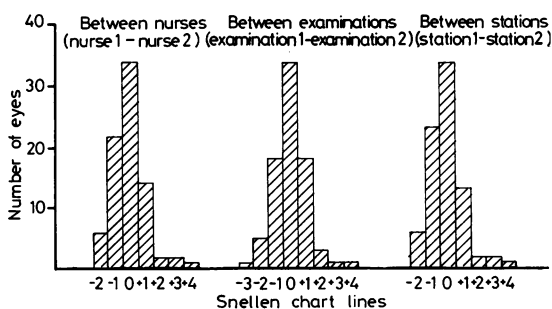


Fig. 2 Frequency distributions of test-retest differences in near visual acuity testing

acuity of 2 print sizes difference or more. Furthermore, this variation was not random, which suggests that one of the nurses was able to encourage patients to attempt smaller print sizes than the other and also that patients were more willing to try on the second test.

The clinical interview data on performance of visual tasks has a higher repeatability than the description of the position of lens opacity or visual acuity results. This may well have an implication for the clinical ophthalmologist. The visual acuity, distance and near, and severity of the lens opacity are the major criteria on which the clinician bases the decision for cataract surgery. The poor repeatability of visual acuity measurement suggests that visual handicap needs to be additionally assessed by careful questioning of the patient's ability to perform everyday visual tasks. The monitoring of patients' progress and the comparison of results with colleagues may be misleading if the poor repeatability of visual acuity measurements is not recognised.

For epidemiologists this small study may serve as yet another warning that there may be dangers in comparing groups of ophthalmic patients by the methods in current use. Studies of the prevalence of ophthalmic conditions and the comparison of groups receiving different treatments need to take

account of the degree of observer variability which may be encountered.

Our thanks are due to the consultant ophthalmologists of the Bristol Eye Hospital and the Royal United Hospital, Bath, for permission to study their patients and in particular to Mr J. C. Dean Hart for reading the paper. We are also grateful to the nursing staff in the outpatient departments of both hospitals and in particular to Nurse M. D. Thomas and Nurse L. R. Gully, who performed the tests of visual acuity.

Professor J. R. T. Colley and Mr A. O. Hughes gave valuable advice on the methodology of the study and the interpretation of the results.

## References

- <sup>1</sup>Wong WW. Indications for cataract surgery. *Arch Ophthalmol* 1978; **96**: 526-28.
- <sup>2</sup>Khan HA, Liebowitz H, Ganley JP, *et al.* Standardising diagnostic procedures. *Am J Ophthalmol* 1975; **79**: 768-75.
- <sup>3</sup>Sanderson HF. Dissertation submitted to the University of London for the degree of MSc Social Medicine. 1978.
- <sup>4</sup>Armitage P. *Statistical Methods in Medical Research*. Oxford: Blackwell Scientific Publications, 1971; 217-68 and 396-7.
- <sup>5</sup>Cochrane AL, Chapman PJ, Oldham PD. Observer errors in taking medical histories. *Lancet* 1951; **1**: 1007-9.
- <sup>6</sup>Hollows FC, Graham PA. Intraocular pressure, glaucoma and glaucoma suspects in a defined population. *Br J Ophthalmol* 1966; **50**: 570-86.
- <sup>7</sup>Brennan ME, Knox EG. The incidence of cataract and its clinical presentation. *Community Health (Bristol)* 1975; **7**: 13-20.